



## FIDUTAM: AI INSIGHT FORUM STATEMENT

Dear Congress and AI Insight Forum Participants,

As the first and largest civil society group in AI whose membership—over 1,500 people ranging from low-income youth to technology policy veterans—represents the broader sepia of humanity, Fidutam recognizes that **addressing current issues with AI affecting diverse stakeholders will feed into a long-term safety agenda.**

We have approached an age of technosolutionism, where AI is an assumed solution to any problem, even when the problem itself is ill-defined. Ignoring the social, political, and cultural dynamics that underlie inherently human challenges will lead to technochauvinism, where digital solutions are considered superior to non-technological ones. **In a technochauvinist climate where AI is not human-aligned, technology will inflame global or existential issues** like climate change, systemic discrimination, political unrest, and weaponization.

Several megacorporations have begun relying on '**digital sweatshops**' for tasks like content moderation, data organization, and other critical procedural work for building or maintaining larger models, raising concerns. At scale, this practice could lead to a problematic dynamic where societal dependence on **AI results in the exploitation of people in low-income or technologically limited communities, compelled to perform mechanical tasks necessary for maintaining these models built by AI companies**, often receiving minimal or no compensation in return.

Another symptom of the AI panacea myth is the concept of **technohacking**. This term, coined by Fidutam, forecasts a scenario **where both adversaries and individual actors gain unregulated access to large AI models**. This access would enable them to create and release potentially hazardous general-purpose systems onto the open web.

While Fidutam as an organization focuses primarily on promoting and developing responsible AI, we also recognize that artificial general intelligence (AGI) may be another tech-facilitated doomsday scenario. This is becoming increasingly plausible as hardware like Graphics processing units (GPUs) improve. **We have therefore appointed members of the Stanford AI Alignment Group to our AI Task Force and experts on AI explainability, such as Daniel Omieza, Ph.D., to our advisory board.** In addition, we consult with existential risk nonprofit the Future of Life Institute on what long-term safeguards could look like in preparation for AGI. Finally, **[we have endorsed nearly 30 bipartisan bills](#) aiming to stave away dangerous or inappropriate use cases of AI.**

**Although AGI's arrival date is still speculative,<sup>1</sup> the principles needed to reel in its potentially catastrophic effects are still vital.** This includes implementing more transparent decision-making architectures in algorithms, introducing human-value reward systems in machine

---

<sup>1</sup> See [How Soon is Now? Predicting the Expected Arrival Date of AGI- Artificial General Intelligence](#) by Rupert Macy-Dare for more information on how an arrival date is being calculated.

learning, maintaining privacy measures, and creating risk evaluation criteria for wholly contextual regulation for different types of AI systems. Fidutam applauds the work of startups like **Humane, Anthropic, and Inflection**, which **are working to bake in trustworthiness and safety into their AI products now**. For example, Humane's Ai Pin allows users to determine what data is collected by the device, whereas Anthropic and Inflection have developed extensive safety measures to reduce the likelihood of their AI chatbots producing misinformation or harmful content.

Beyond industry trends, there are academic factions forming in the AI space. **Effective Altruism (EA)** has become the vogue of the AI movement, promoting the use of evidence and rational analysis to optimize positive impact. This approach frequently applies the idea of **longtermism**, which emphasizes the moral significance of shaping the long-term future positively. More recently, **EA has appended itself to long-term AI safety**, introducing a new political voice.<sup>2</sup>

Adherents of Effective Altruism in AI, comprising some of the most influential and affluent names in tech, focus their actions and resources on addressing the long-term, existential risks linked to artificial intelligence. This not only diverts support away from groups addressing algorithmic injustice, but it can also overshadow the concerns raised by underrepresented communities. The trend therefore **risks allowing companies, governments, and developers to overlook or persist in discriminatory practices through AI, often without public accountability or legal consequences**.

It also cannot be ignored that **EA finds its roots and leadership with figures** like Nick Bostrom, Nick Cooney, and Stijn Bruers, all of **whom have had well-documented ties to racism, racial prejudice, or the undermining of systemic racism**.<sup>3</sup> The movement has yet to fully distance itself from this alarming literature and rhetoric, despite permeating Silicon Valley and the broader AI community. We must all be wary that **potentially discriminatory schools of thought could play a significant role in how AI evolves**.

Due to EA's unfortunate ties and, in some cases, extremism, there has been an abundance of infighting between long-term AI risk and near-term AI justice groups on what timeline and approach should take priority. **This conflict has presented a false choice to Congress and other regulators that must be rejected**: it is possible to tackle both near-term and long-term risks of AI, and they are often interlaced; solving issues like biased predictive policing, misinformation, or unrestricted mass surveillance will invariably support the development of human-aligned AI.

Doomsday scenarios shouldn't be limited to AI's omnipotence or potential to create bioweapons. If digital sweatshops, algorithmic injustices, and unaligned AI are not addressed, we may find our citizens and democracy in an existential state of collapse.<sup>4</sup> To prevent this, **dialogue and co-development between industry, government, and civilian stakeholders without obtrusive secrecy is necessary for both current and frontier AI**.

Fidutam is ready to represent the crystallization of such collaboration and hopes that more leaders will join us in the multistakeholder pursuit of responsible AI. Only together can we engineer a future where AI works in the interest of all people.

---

<sup>2</sup> This is not intended to attack or condemn the [Effective Altruism](#) organization or their extensive philanthropic work. We are spotlighting problematic associations and their relation to AI.

<sup>3</sup> [Anti-Blackness and the Effective Altruist](#) by Christopher Sebastian and Nick Bostrom's own "[Apology for an Old Email](#)" cover this in more detail.

<sup>4</sup> See our [Guarding Against AI Doomsday Scenarios + AI in National Security & Risk Management](#) Memo for more information on risks posed by big AI.